“Zeppelin”: an SoC for Multi-chip Architectures

Noah Beck[1], Sean White[1], Milam Paraschou[2], Samuel Naffziger[2]

[1]AMD, Boxborough, [2]AMD, Fort Collins

Presented at ISSCC 2018

AMD

Outline

- Design Goals for the System-on-a-Chip codenamed “Zeppelin”
- SoC Architecture
- Core Complex codenamed “Zen”
- AMD Infinity Fabric (IF)
- I/O Capabilities, I/O muxing
- Floorplan and Packaging
- Results

“Zeppelin” SoC Goals

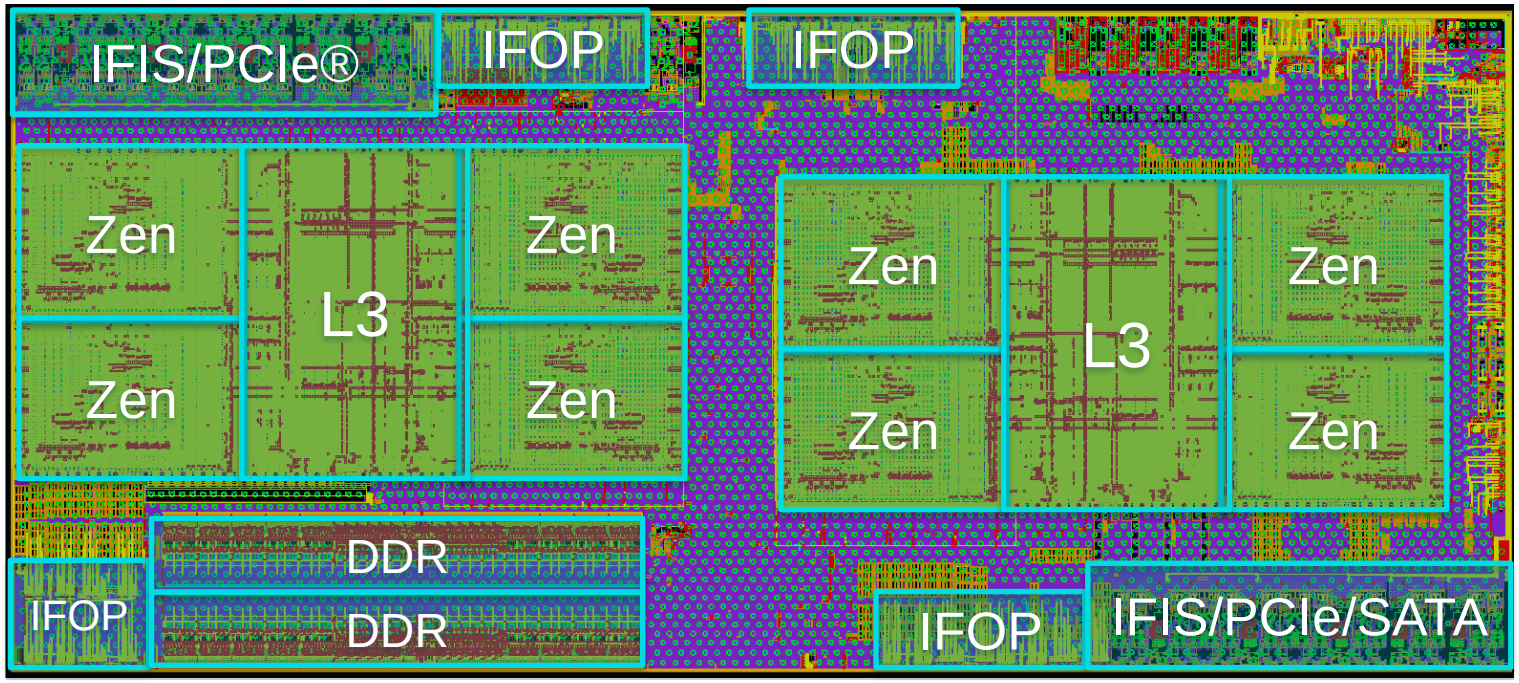
Design a System-on-a-Chip Solution for scalability across the Server market

---

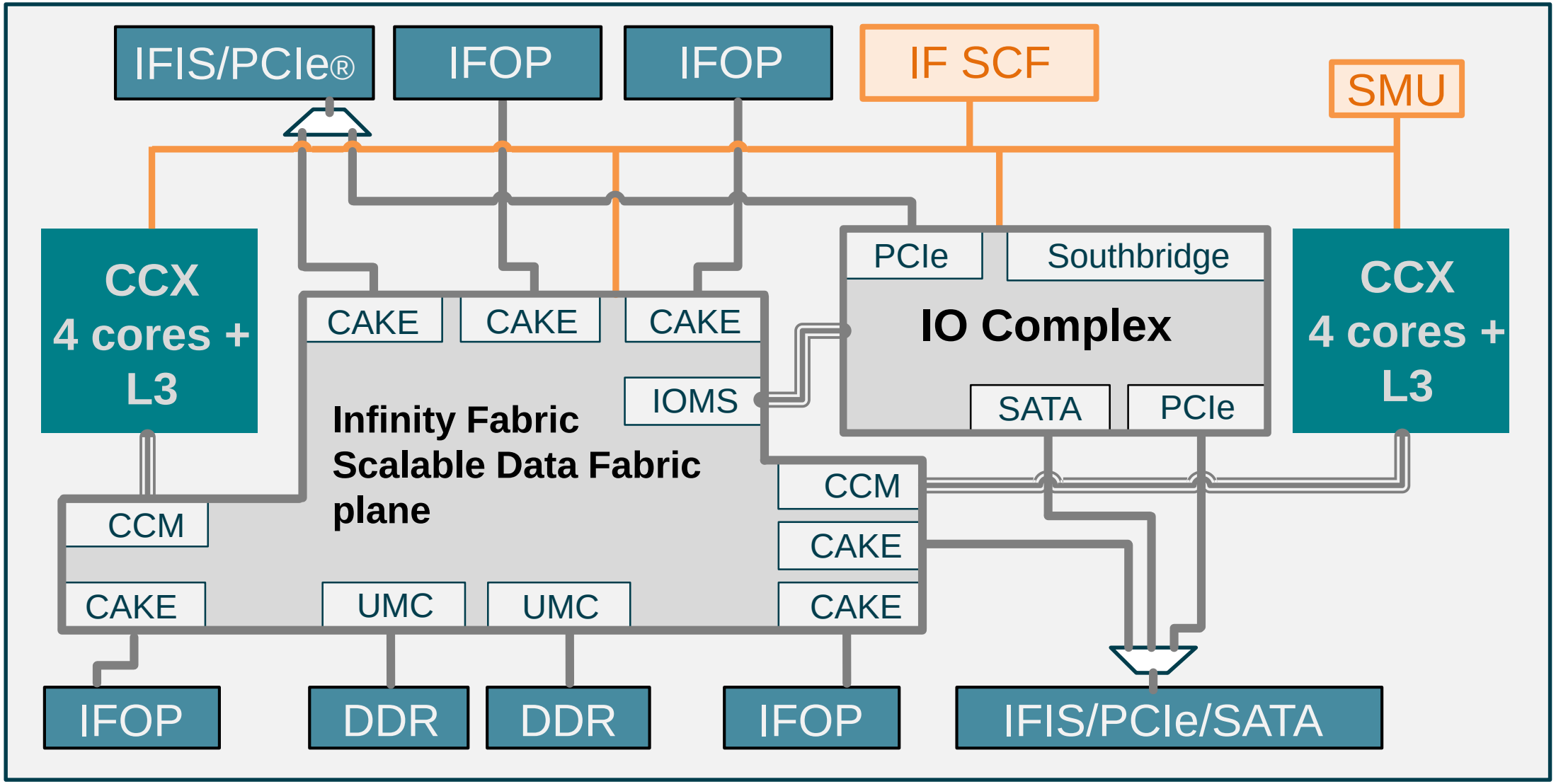
- **4-die multi-chip module (MCM) for Server in new infrastructure**
- **Same SoC suitable for High-End Desktop**
  - 1-die Desktop in existing AM4 infrastructure
  - 2-die MCM High-End Desktop in new infrastructure

“Zeppelin” Die Functional Overview

- **Compute**
  - 8 “Zen” x86 cores
  - 4MB total L2 cache
  - 16 MB total L3 cache
- **Memory**
  - 2 channel DDR4 with ECC
  - 2 DIMMs/channel and up to 256GB/channel
- **Integrated I/O**
  - Coherent and control Infinity Fabric links
  - 32 lanes high-speed SERDES
  - 4 USB3.1 Gen1 ports
  - Server Controller Hub (SPI, LPC, UART, I2C, RTC, SMBus)

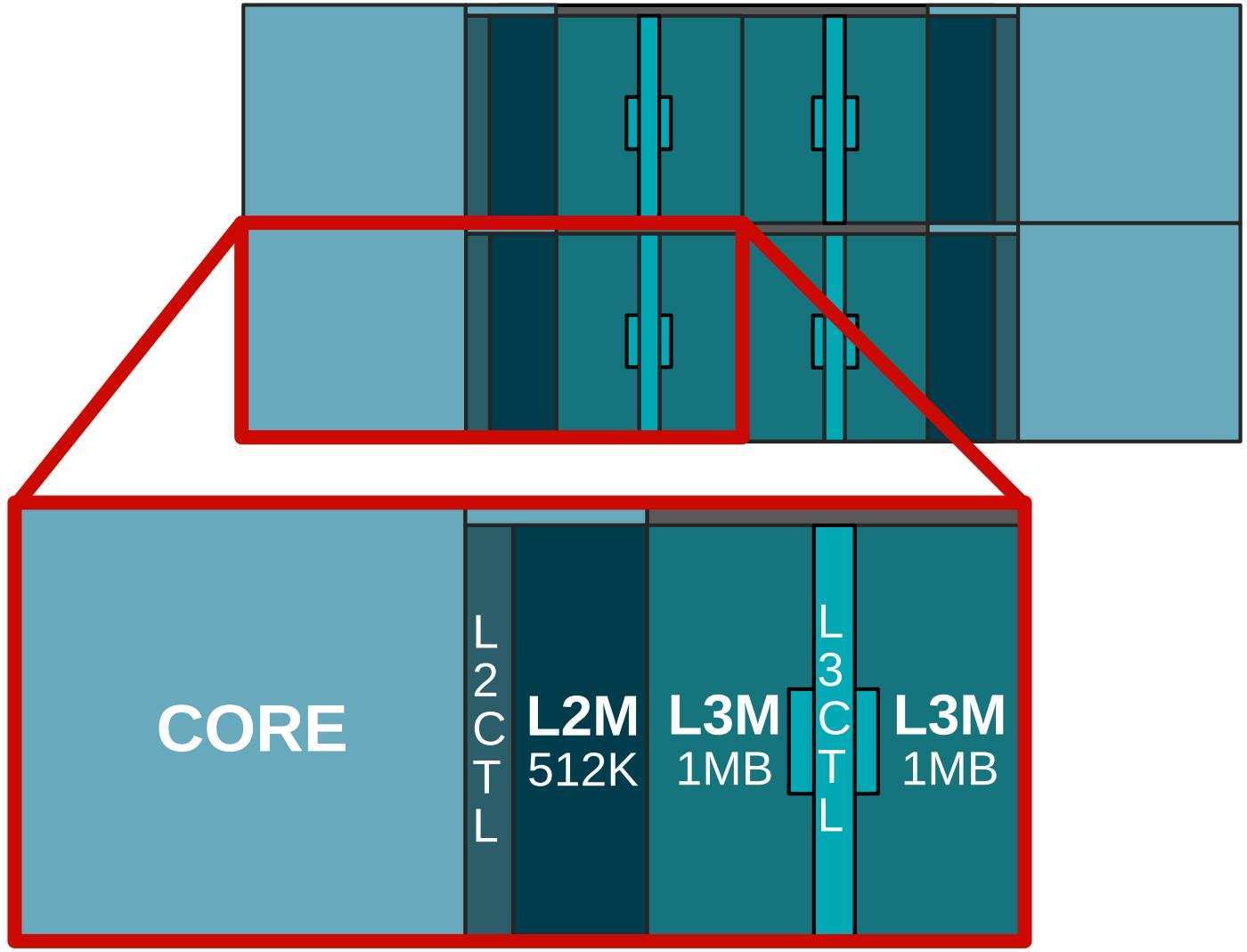


Chip Architecture



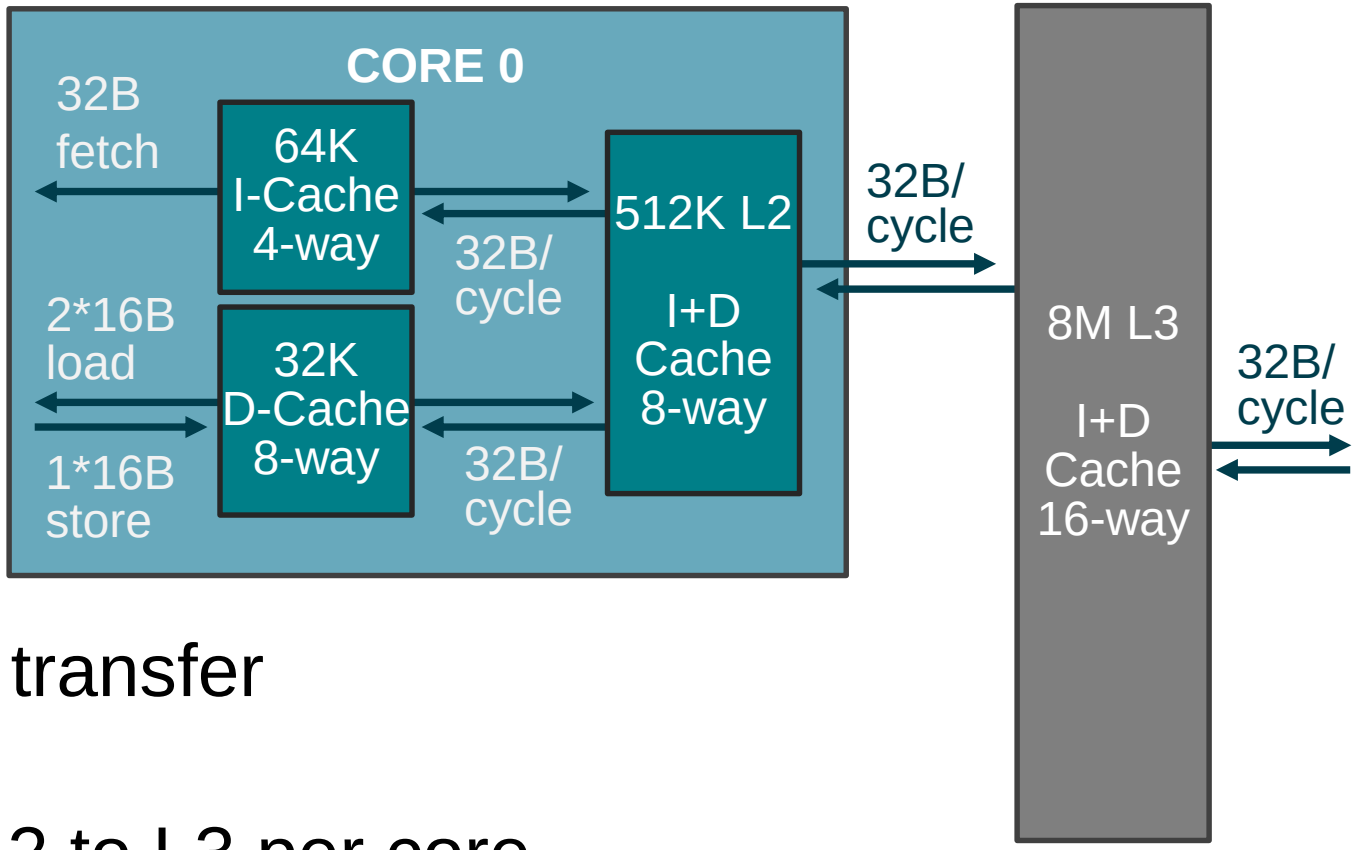
CCX: CPU Complex

- 4 cores with L1/L2 caches, plus shared L3 cache
- “Zen” core described in [Singh ISSCC17]
  - L1 Instruction Cache 64KB, 4-way associative
  - L1 Data Cache 32KB, 8-way associative
  - L2 Cache 512KB, 8-way associative
  - 2 threads per core
- L3 cache 8MB, 16-way associative, shared by all four cores

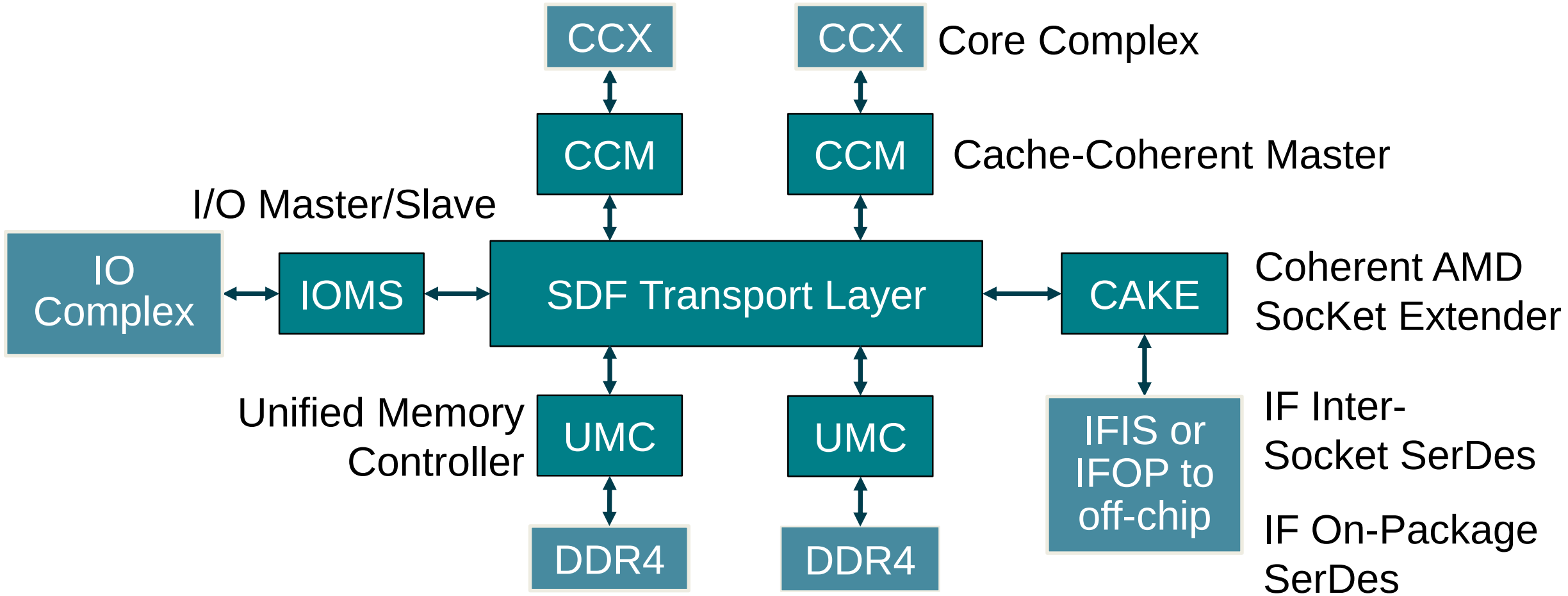


“Zen” Cache hierarchy

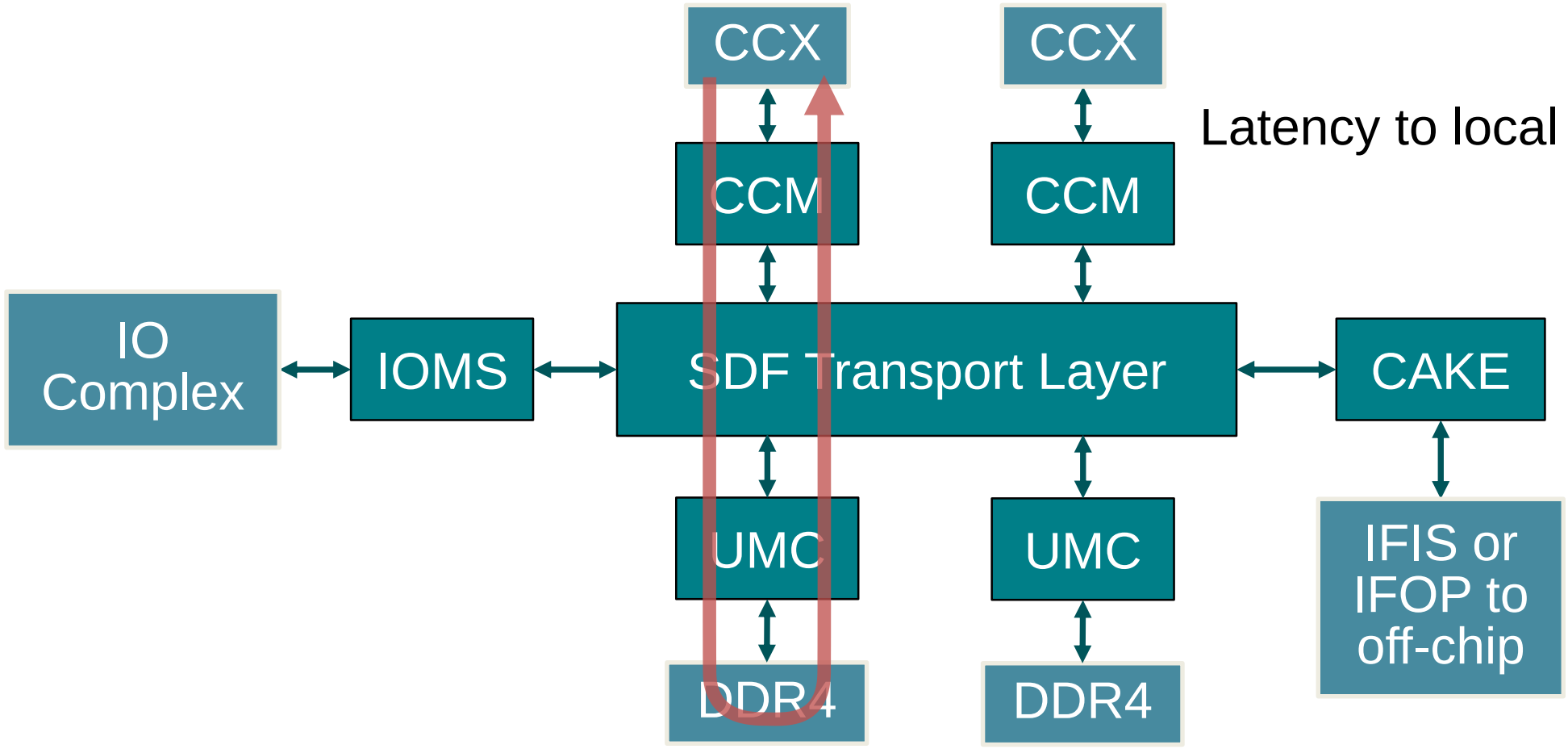
- Fast private L2 cache, 12 cycles
- Fast shared L3 cache, 35 cycles
- L3 filled from L2 victims of all four cores
- L2 tags duplicated in L3 for probe filtering and fast cache transfer
- Multiple smart prefetchers
- 50 outstanding misses from L2 to L3 per core
- 96 outstanding misses from L3 to memory



AMD Infinity Fabric: Scalable Data Fabric



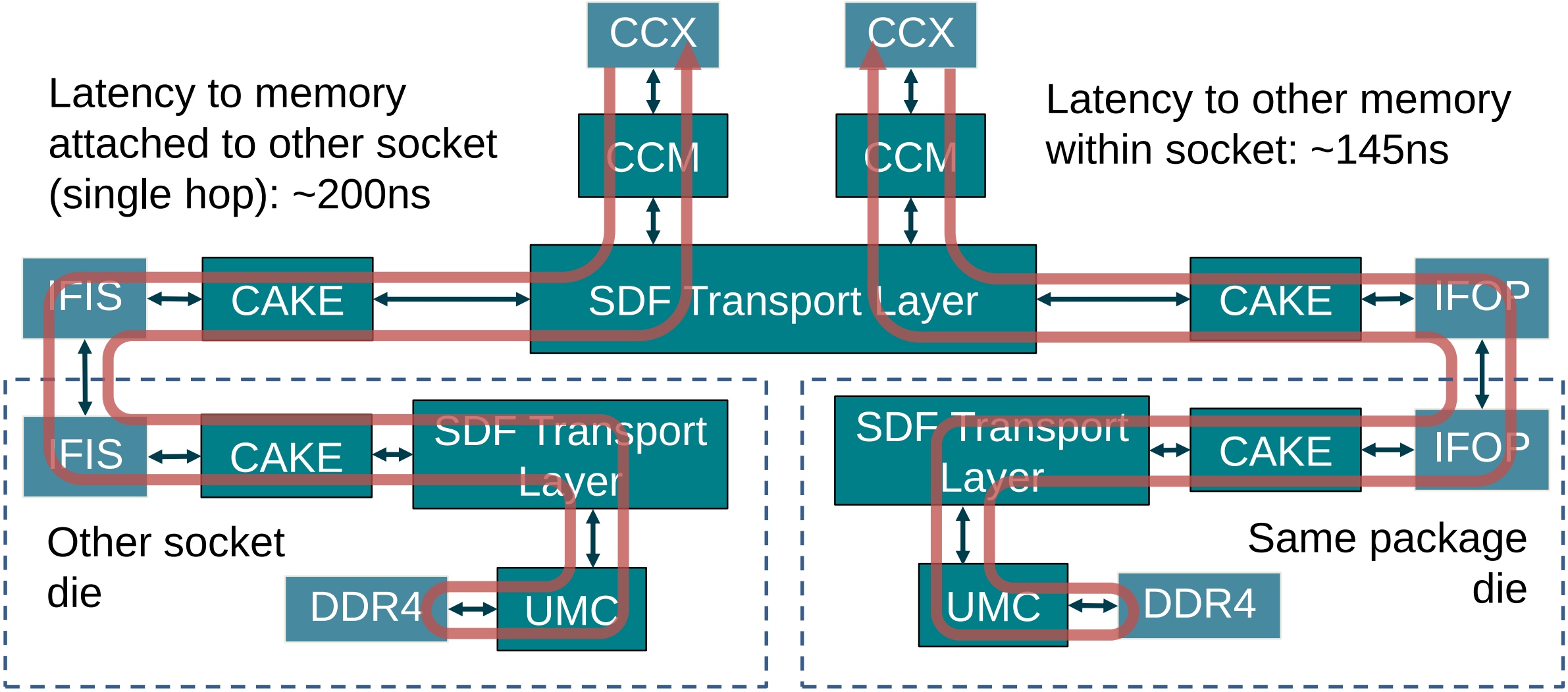
SDF Local Memory Access



Latency to local memory: ~90ns

* See Endnotes for additional system configuration details

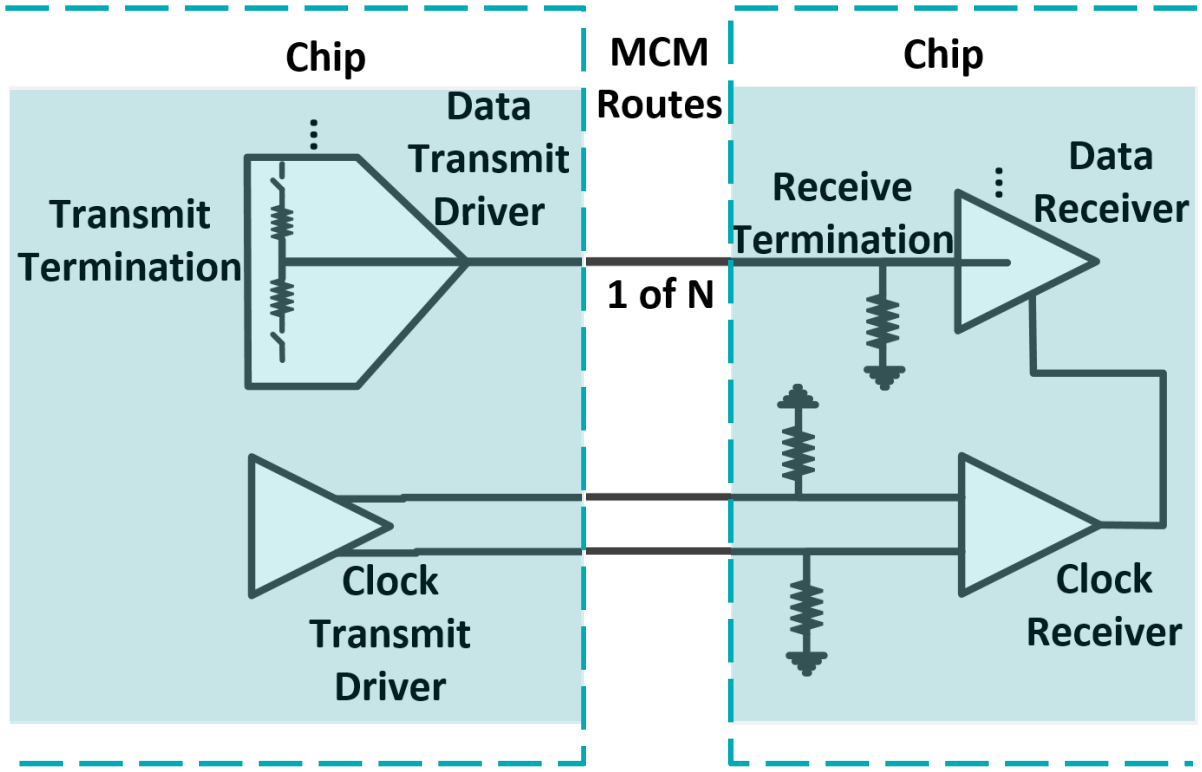
SDF Die-to-Die Memory Accesses



* See Endnotes for additional system configuration details

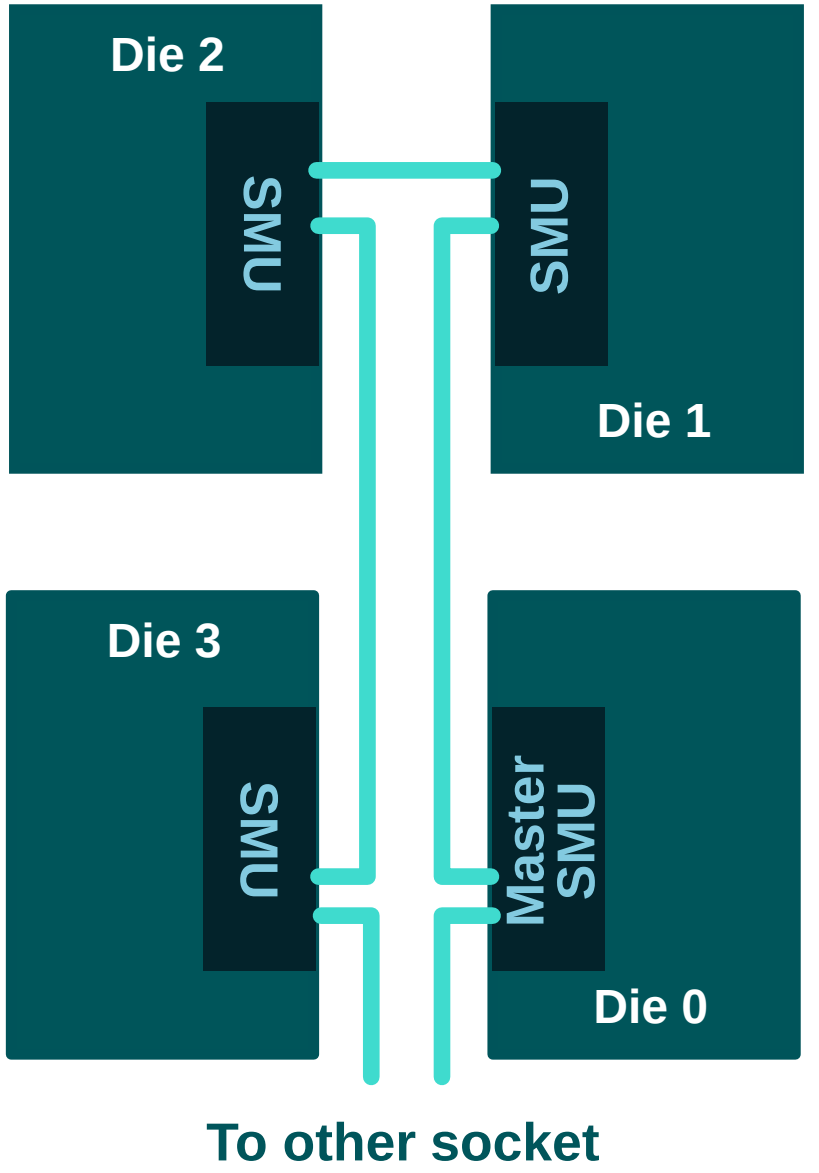
2pJ/bit IFOP SerDes

- Low-swing, single-ended data for ~50% of power of an equivalent differential driver
- Zero power driver state during logic 0 transmit
  - Transmit/receive impedance termination to ground while driver pullup is disabled
  - Also applied during link idle
- Data bit inversion encoding saving 10% average power per bit



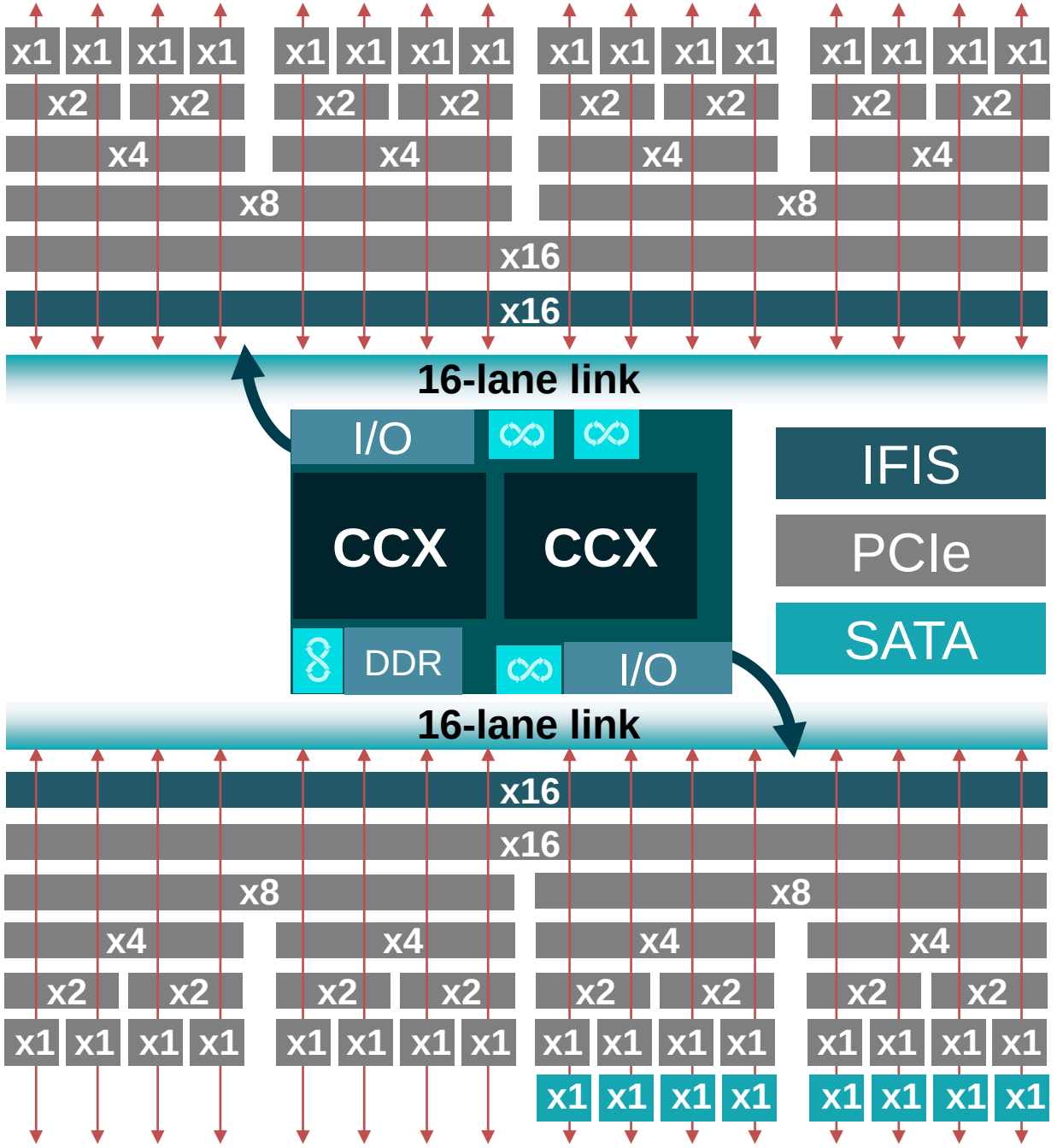
Hierarchical Power Management

- System Management Unit (SMU) uses IF Scalable Control Fabric (SCF) plane
- SCF: single-lane IFIS SerDes link for chip-to-chip or socket-to-socket
- SMU calculation hierarchy for voltage level control, C-State Boost, thermal management, electrical design current management
  - Local chip SMU fast loop
  - Master chip SMU slower loop



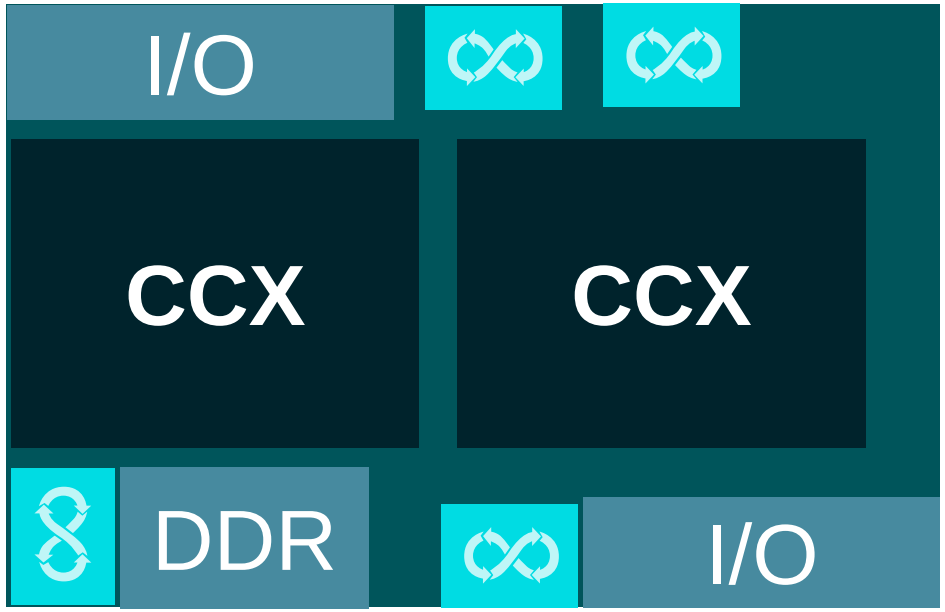
IO Subsystem & Muxing

- 32 lanes multi-protocol I/O
  - PCIe, IFIS: two 16-lane links
  - PCIe link bifurcation:
    max 8 devices per 16-lane link
  - SATA: 8 lanes of bottom link
- Supports multiple market segments
- Muxing support adds
  <1 channel clock latency to IFIS



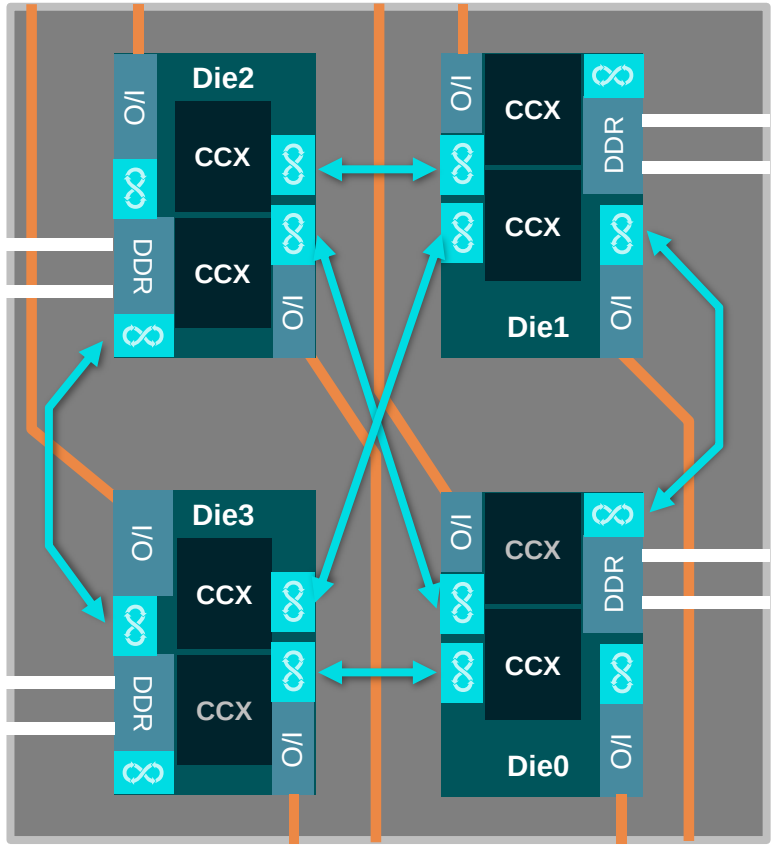
Chip Floorplanning for Package

- DDR placement on one die edge
- Chips in 4-die MCM rotated 180°, DDR facing package left/right edges
- Package-top Infinity Fabric pinout requires diagonal placement of IFIS
- 4th IFOP enables routing of high-speed I/O in only four package substrate
  layers

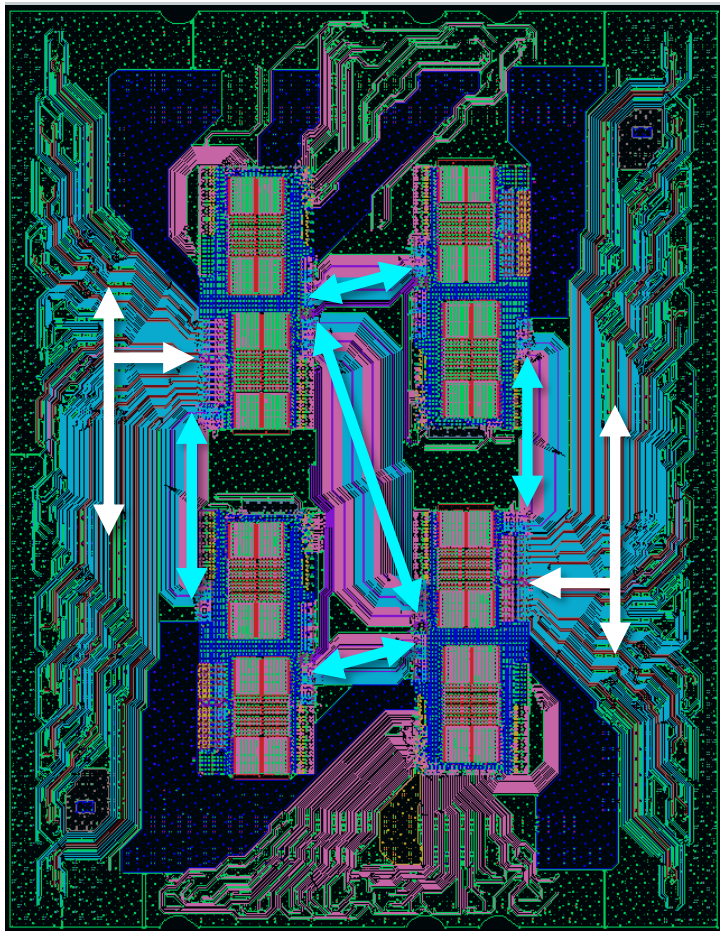


DDR+IFOP Package Routing

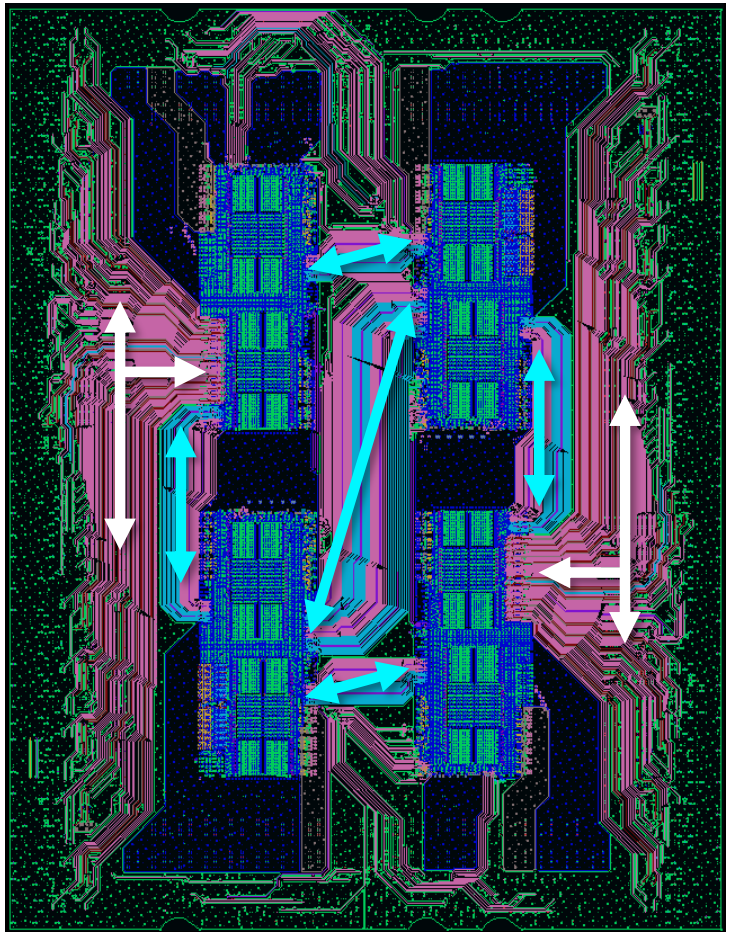
- Vertical and Horizontal IFOP: 2 layers each
- Diagonal IFOP: 1 layer each
- DDR channel: 1 layer each



Layer A

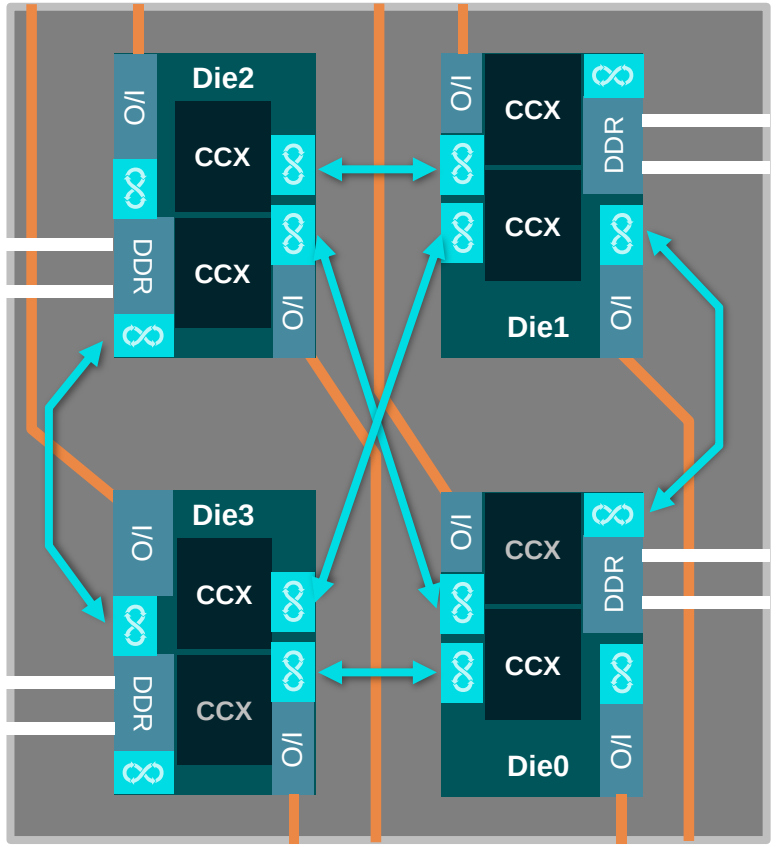


Layer B

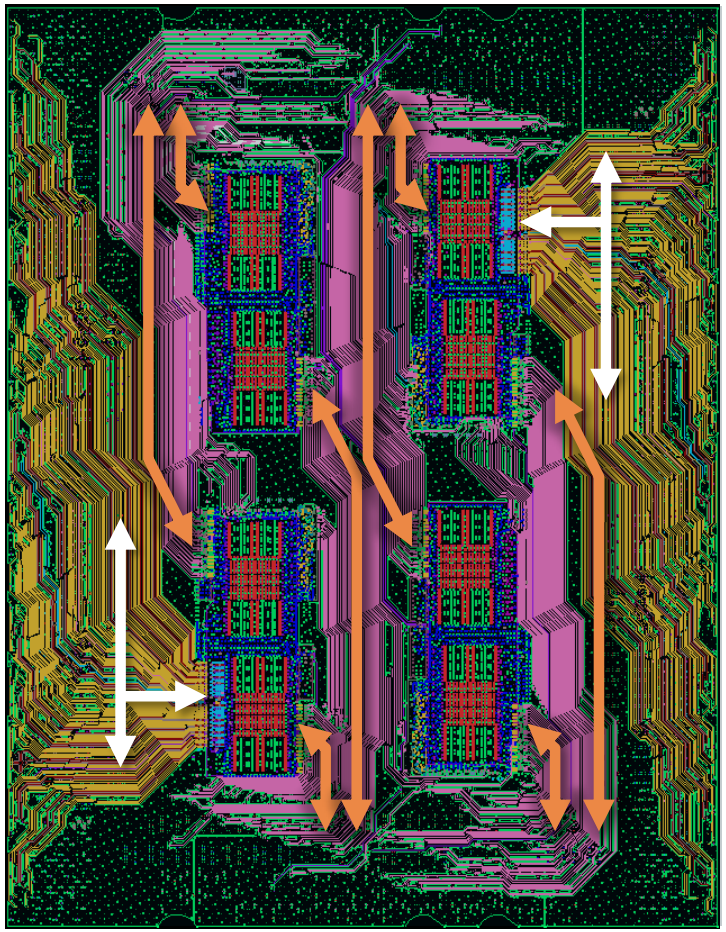


DDR+IFIS Package Routing

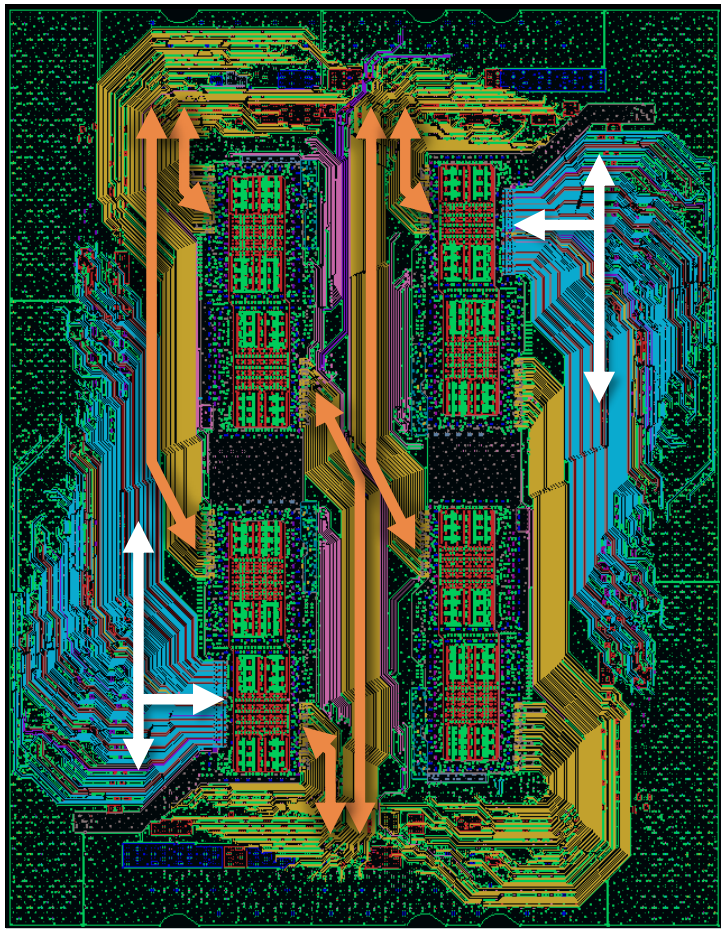
- DDR channel: 1 layer each
- IFIS links: 2 layers each



Layer C



Layer D

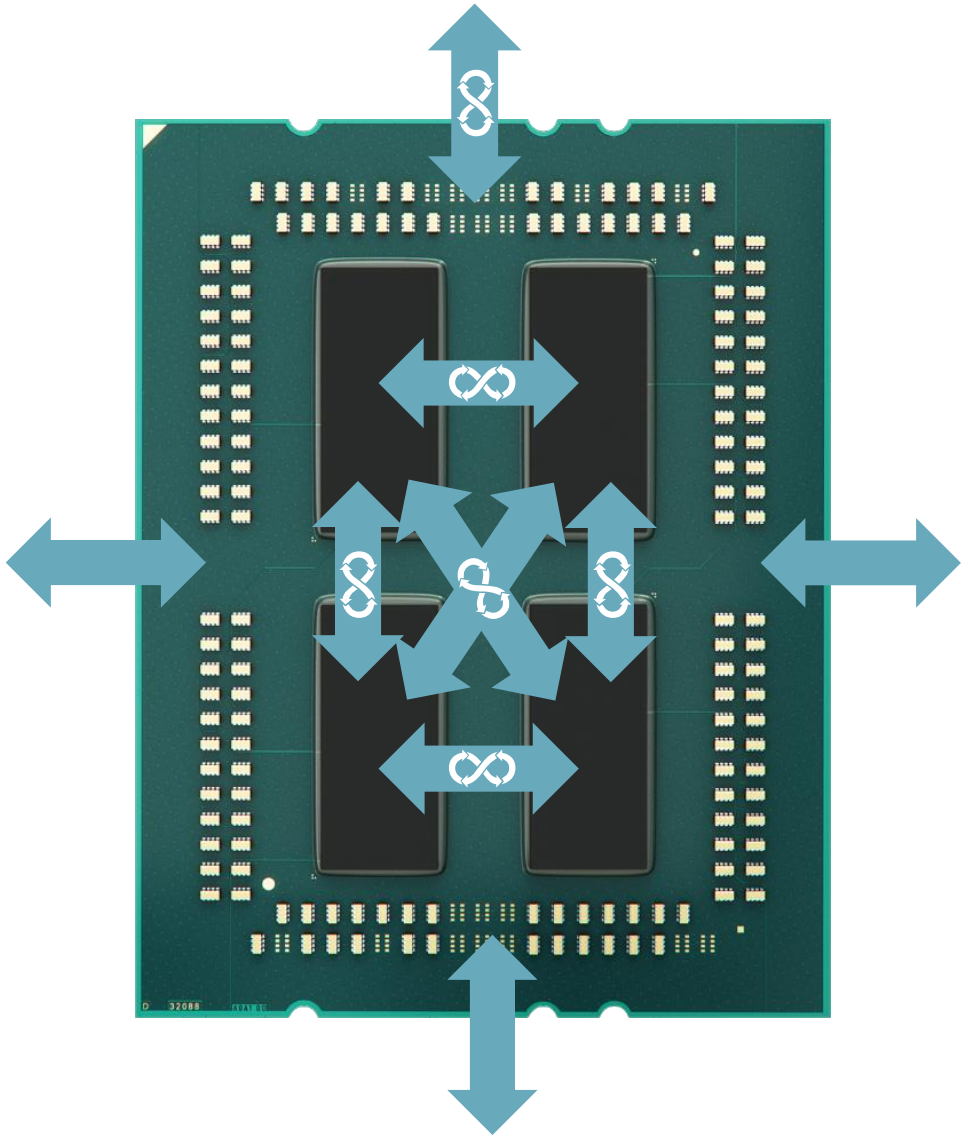


MCM Versus Single-Chip Design

- 4-die MCM package: 852mm^2 of silicon (4 * 213mm^2)
- Large single-chip design:
  - ~10% area savings: 777mm^2 (near reticle size limit)
  - Manufacturing/test cost: ~40% higher
  - Full 32-core yield: ~17% lower
  - Full 32-core cost: ~70% higher
- High-yielding multi-chip assembly process
  - Achievable based on internal production data
  - Die frequency matching using on-die frequency sensors

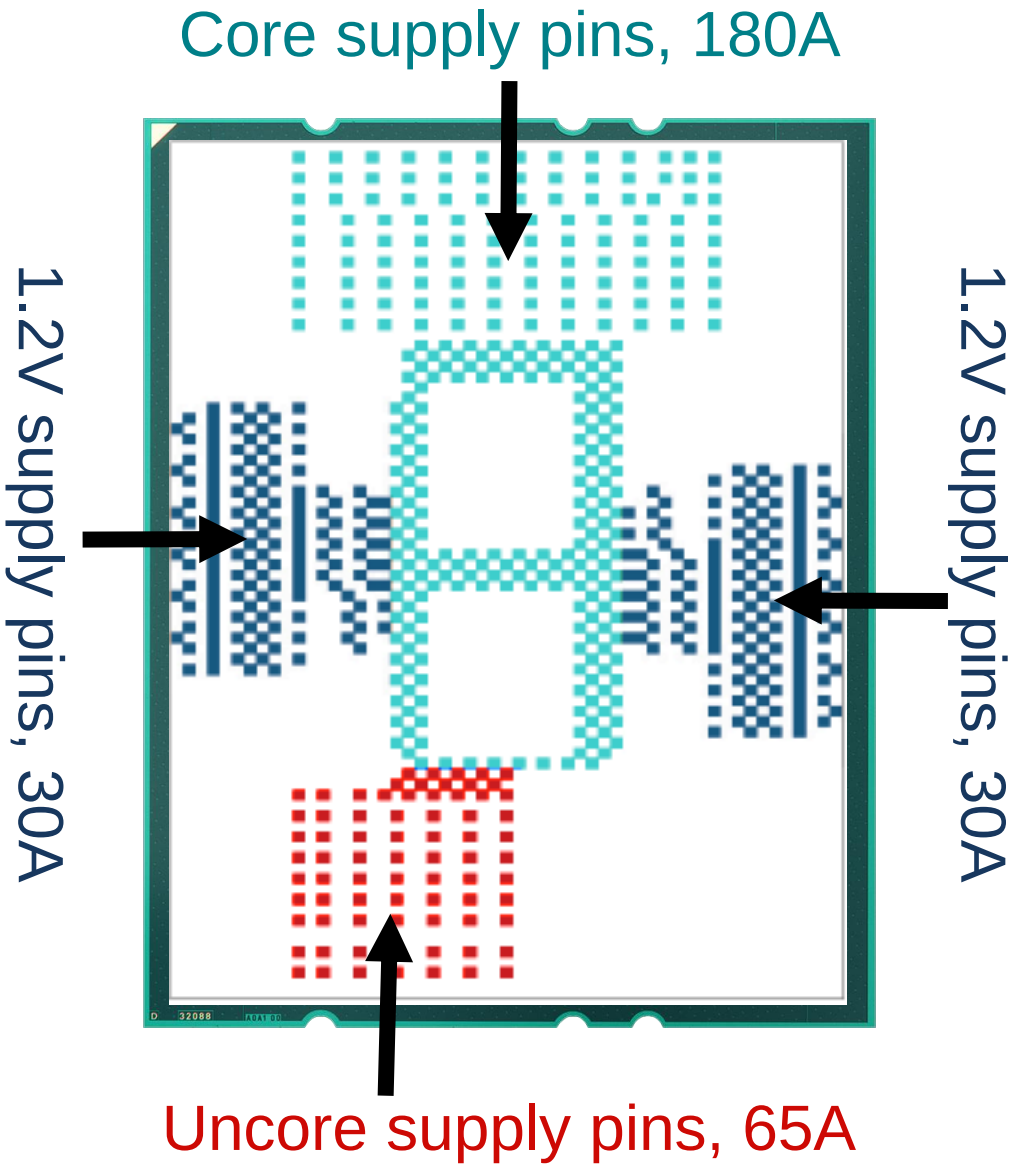
MCM Package Achievements

- 4094 total LGA pins
- 58mm x 75mm organic substrate
- 534 IF high-speed chip-to-chip nets
  - Over 256GB/s total in-package bandwidth
- 1760 high-speed pins
  - Over 450GB/s total off-package bandwidth



More MCM Package Achievements

- ~300µF of on-package cap
- ~300A current
- Up to 200W TDP



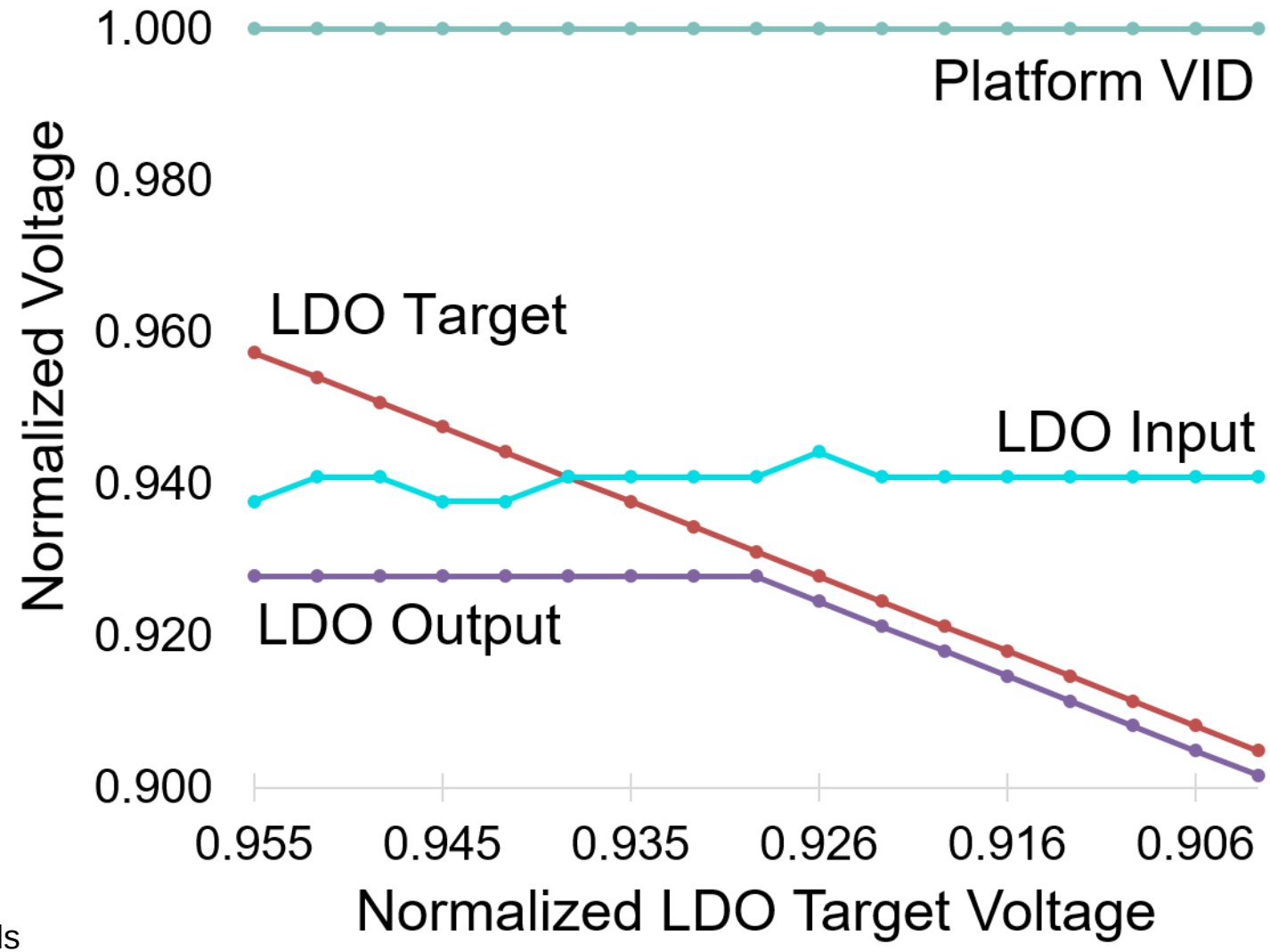
MCM Core Voltage Variation

- Per-core measurements shown
  - +/-25mV accuracy with max power workload
- Per-core ring oscillators
  - Calibrated for temperature and voltage
  - Min/max voltage sampled 470M/s
- Static differences compensated by per-core LDOs
- Dynamic differences mitigated by clock stretcher, DPM states

* See Endnotes for additional system configuration details

Core Voltage Measurements

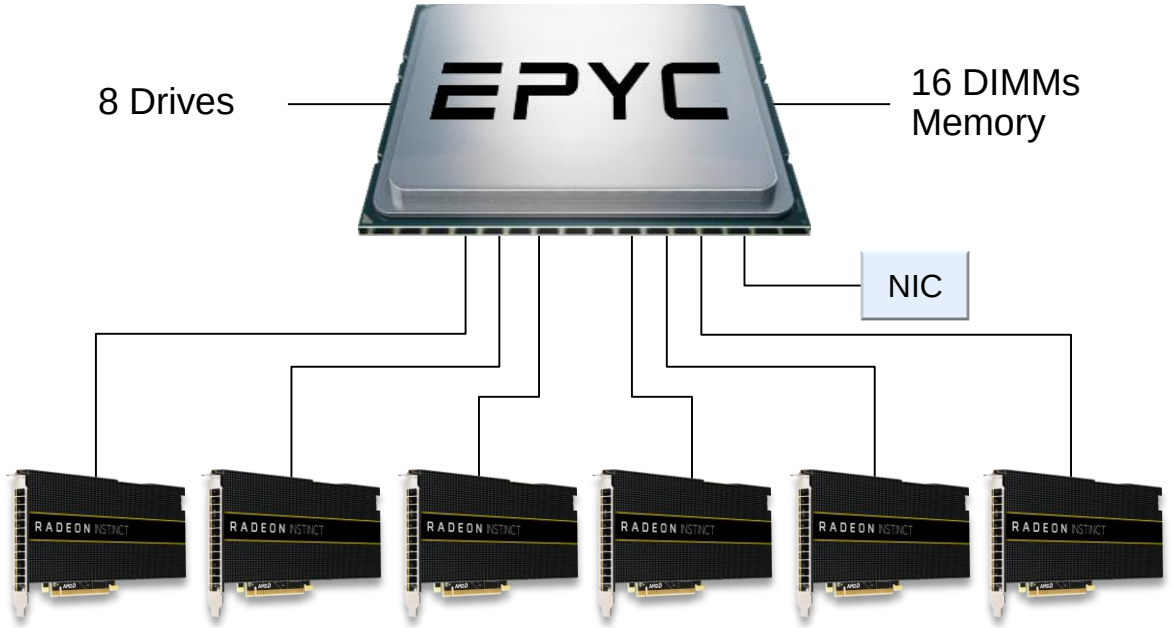
- Measured data shows excellent tracking of per-core voltage from the digital LDO with mV-accurate target voltage
- Power savings through per-core voltage optimization

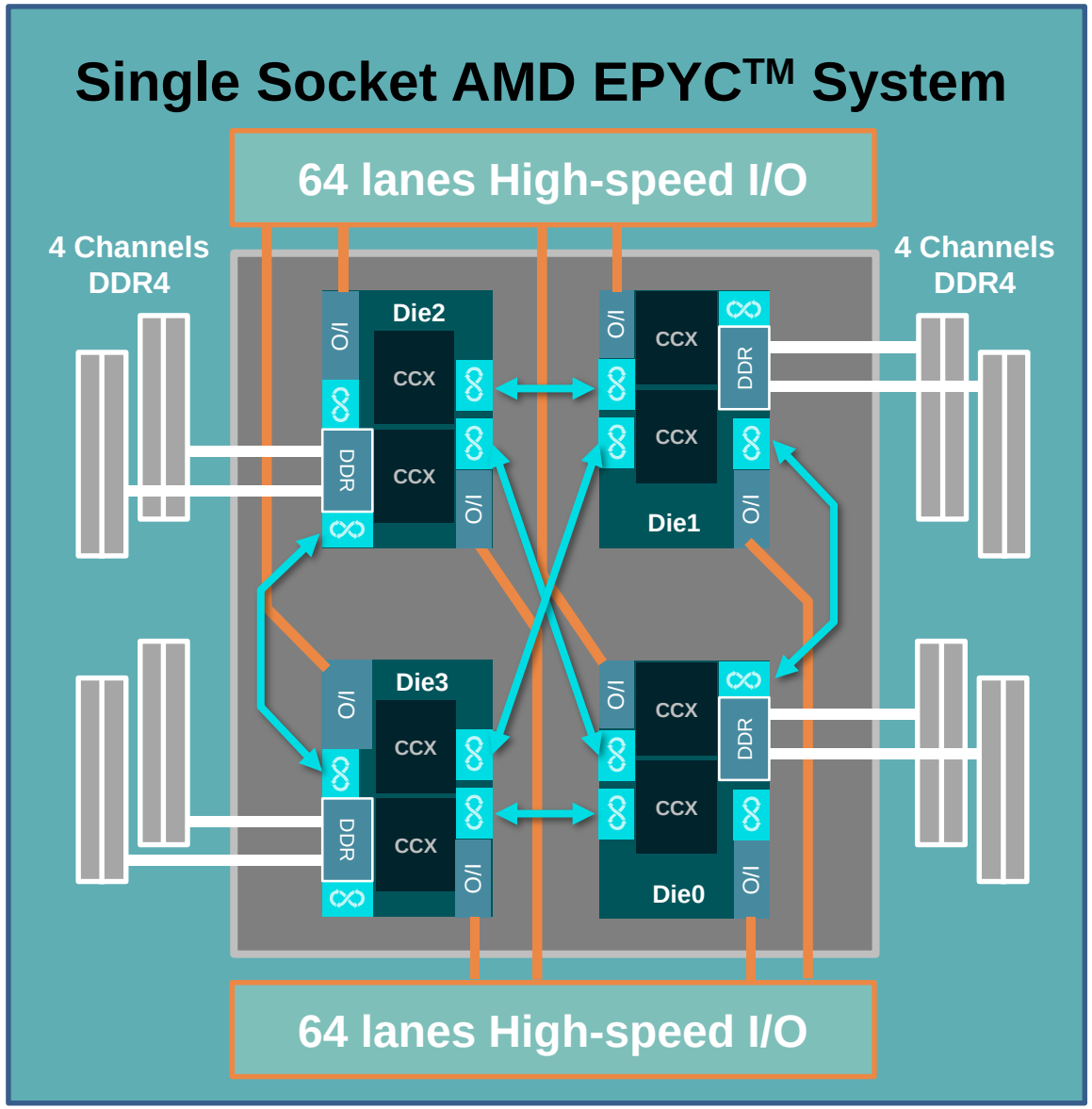


* See Endnotes for additional system configuration details

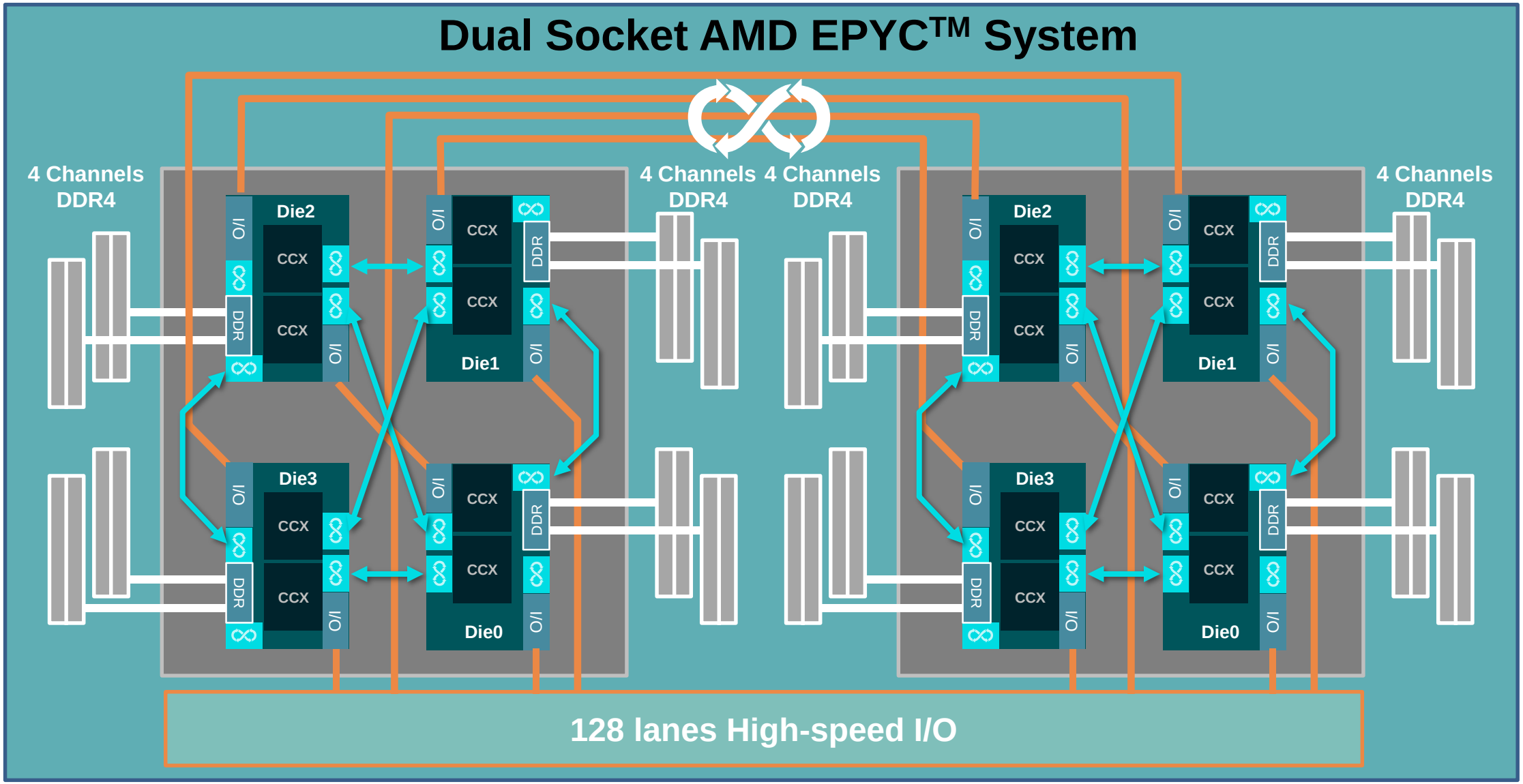
4-Chip EPYC Package

- 128 lanes can be used as PCIe
  - Attach **six** 16-lane accelerator cards to a single socket
- 8 DDR4 channels



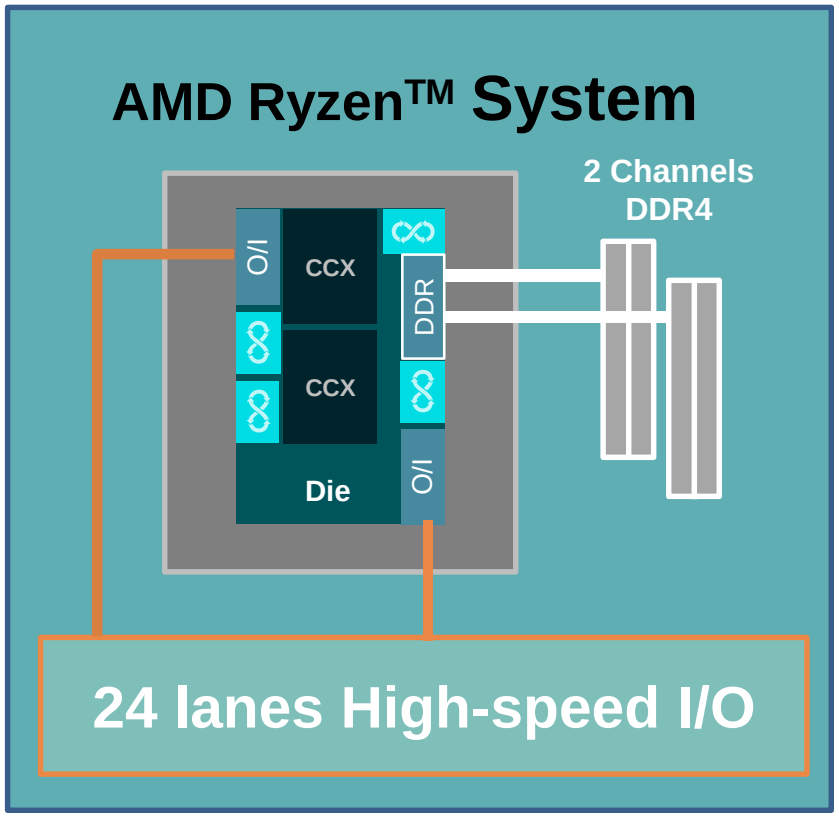


Dual 4-Chip EPYC Packages



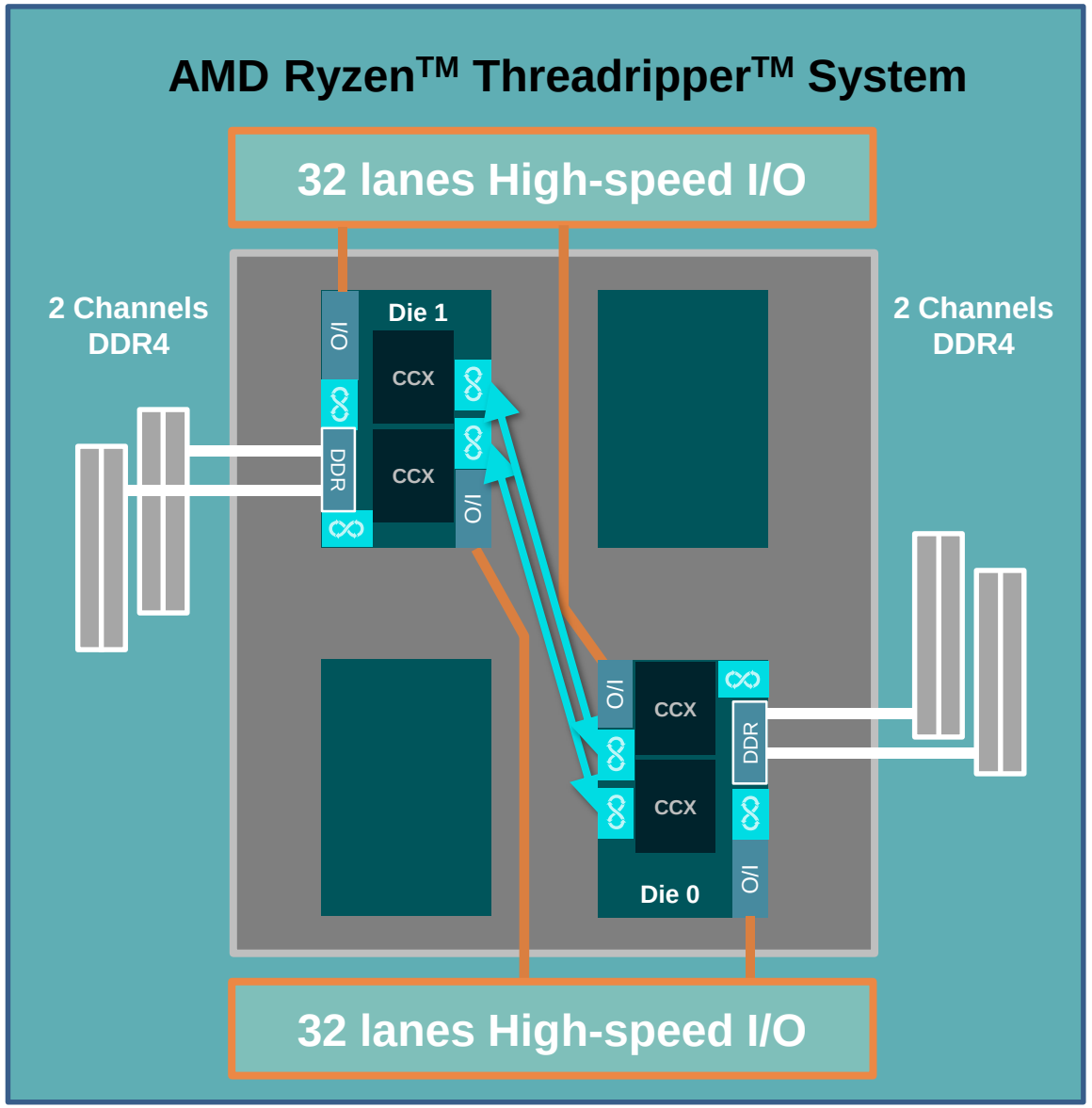
Single Chip AM4 Package

- Socket compatible with other AMD SoCs for desktop market
- 8 cores / 16 threads
- 2 DDR4 channels
- 24 PCIe Gen3 lanes
- Up to 95W TDP



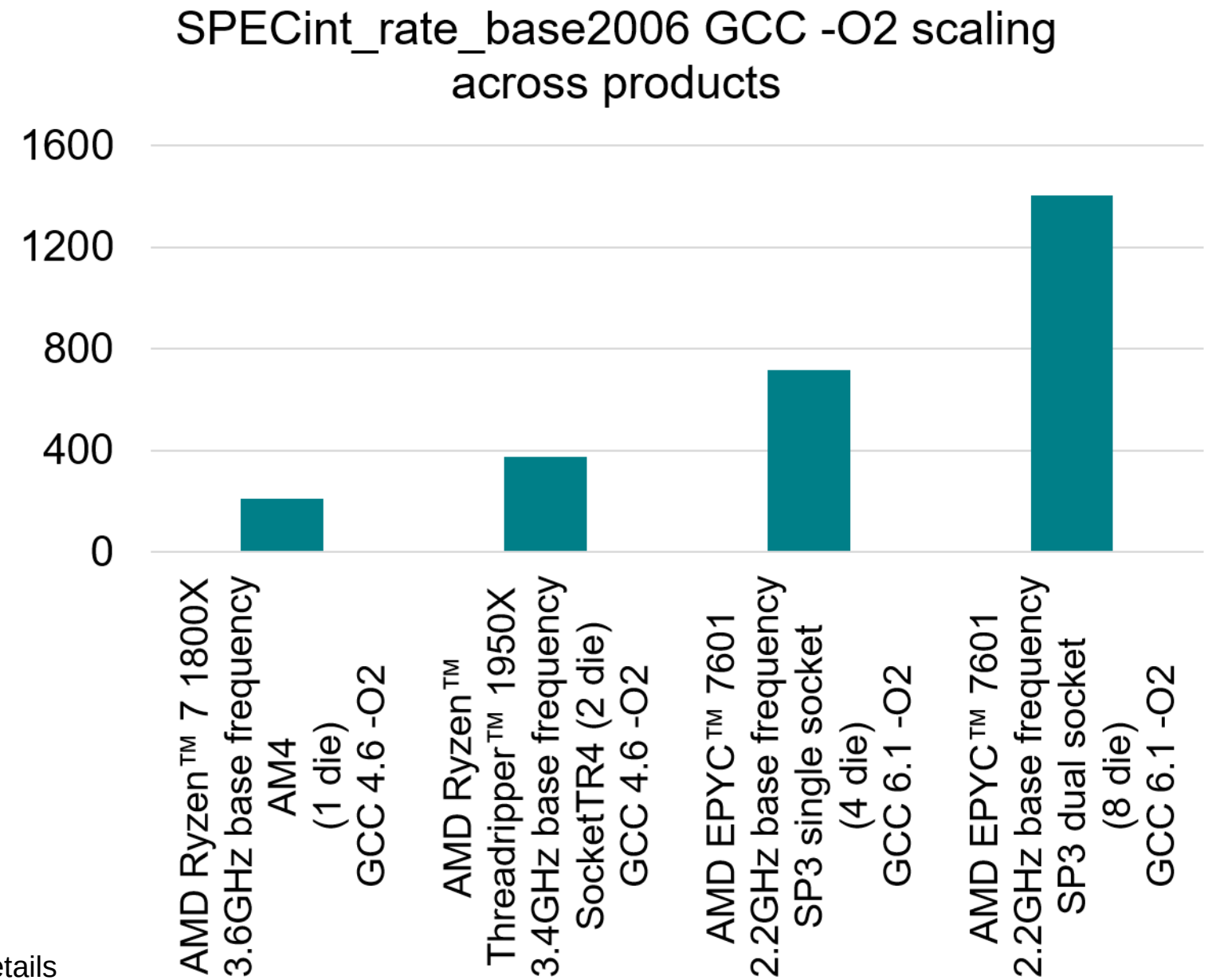
2-Chip sTR4 Package

- Socket defined for “Zeppelin” SoC and compatible with future designs
- 16 cores / 32 threads
- 4 DDR4 channels
- 64 PCIe Gen3 lanes



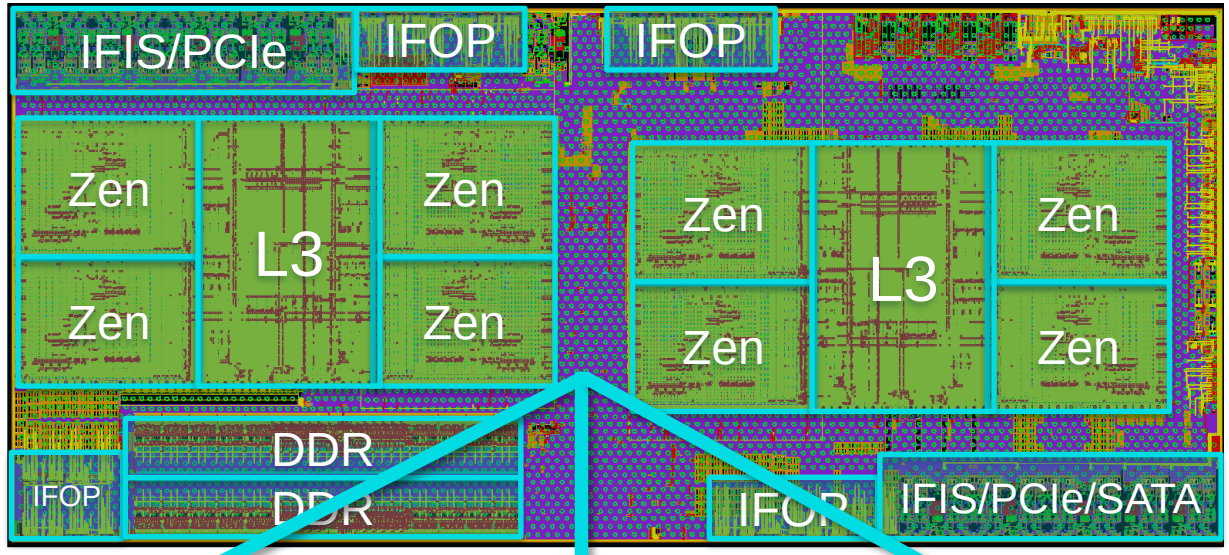
Benchmark results

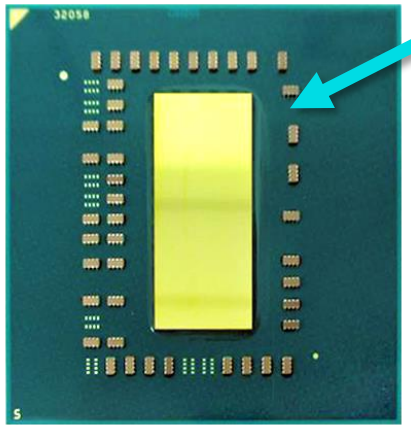
- Scalable performance from single-chip up to 8-chip 2-socket configuration

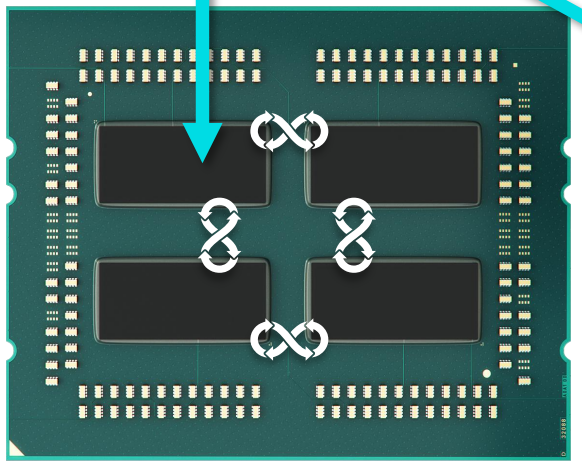


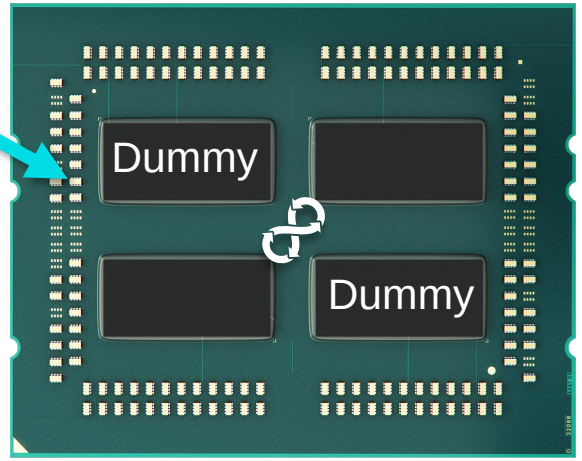
* See Endnotes for additional system configuration details

An SoC for Multi-chip Architectures









Mainstream Desktop

Performance Server

High-End Desktop

Acknowledgment

- We would like to thank our talented AMD design teams across Austin, Bangalore, Boston, Fort Collins, Hyderabad, Markham, Santa Clara, and Shanghai, who contributed on “Zen” and “Zeppelin”
- Please check out our demo tonight

Endnotes

Slides 9, 10:
Latencies assume 2.4GHz CPU core frequency and 1R DDR4-2667 19-19-19 RDIMM; Memory, IFIS, IFOP latencies are dependent on DRAM clock; Memory latencies include testing overhead (including DRAM refresh).

Slides 20, 21:

Power measurements taken from a SP3 Diesel non-DAP AMD evaluation system, with EPYC rev B1 parts, BIOS revision WDL7405N, Windows Server 2016, running a Max Power pattern at 2.5GHz core frequency

Slide 26:

AMD Ryzen™ 7 1800X CPU scored 211, using estimated scores based on testing performed in AMD Internal Labs as of 30 March 2017. System config: Ryzen™ 7 1800X: AMD Myrtle-SM with 95W R7 1800X, 32GB DDR4-2667 RAM, Crucial CT256M550SSD, Ubuntu 15.10, GCC –O2 v4.6 compiler suite.

AMD Ryzen™ Threadripper™ 1950X CPU scored 375, using estimated scores based on testing performed in AMD Internal Labs as of 7 September 2017. System config: Ryzen™ Threadripper™ 1950X: AMD Whitehaven-DAP with 180W TR 1950X, 64GB DDR4-2667 RAM, CT256M4SSD disk, Ubuntu 15.10, GCC –O2 v4.6 compiler suite.

AMD EPYC™ 7601 CPU scored 702 in a 1-socket using estimated scores based on internal AMD testing as of 6 June 2017. 1 x EPYC™ 7601 CPU in HPE Cloudline CL3150, Ubuntu 16.04, GCC -O2 v6.3 compiler suite, 256 GB (8 x 32 GB 2Rx4 PC4-2666) memory, 1 x 500 GB SSD

AMD EPYC™ 7601 scored 1390 in a 2-socket system using estimated scores based on internal AMD testing as of 6 June 2017. 2 x EPYC™ 7601 CPU in Supermicro AS-1123US-TR4, Ubuntu 16.04, GCC -O2 v6.3 compiler suite, 512 GB (16 x 32GB 2Rx4 PC4-2666 running at 2400) memory, 1 x 500 GB SSD.